

High accuracy deep learning wavefront sensing under high-order turbulence

Dongming Liu^{1,2} , Hui Liu^{1*} , Zhenyu Jin¹

¹Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216, China

²University of Chinese Academy of Sciences, Beijing 100049, China

*Correspondence: liuhui@ynao.ac.cn

Received: September 11, 2024; Accepted: September 24, 2024; Published Online: November 2, 2024; <https://doi.org/10.61977/ati2024052>; <https://cstr.cn/32083.14.ati2024052>

© 2024 Editorial Office of Astronomical Techniques and Instruments, Yunnan Observatories, Chinese Academy of Sciences. This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>)

Citation: Liu, D. M., Liu, H., Jin, Z. Y. 2024. High accuracy deep learning wavefront sensing under high-order turbulence. *Astronomical Techniques and Instruments*, 1(6): 316–324. <https://doi.org/10.61977/ati2024052>.

Abstract: We explore an end-to-end wavefront sensing approach based on deep learning, which aims to deal with the high-order turbulence and the discontinuous aberration caused by optical system obstructions commonly encountered in real-world ground-based telescope observations. We have considered factors such as the entrance pupil wavefront containing high-order turbulence and discontinuous aberrations due to obstruction by the secondary mirror and spider, realistically simulating the observation conditions of ground-based telescopes. By comparing with the Marechal criterion (0.075λ), we validate the effectiveness of the proposed approach. Experimental results show that the deep learning wavefront sensing approach can correct the distorted wavefront affect by high-order turbulence to close to the diffraction limit. We also analyze the limitations of this approach, using the direct zonal phase output method, where the residual wavefront stems from the fitting error. Furthermore, we have explored the wavefront reconstruction accuracy of different noise intensities and the central obstruction ratios. Within a noise intensity range of 1%–1.9%, the root mean square error (RMSE) of the residual wavefront is less than Marechal criterion. In the range of central obstruction ratios from 0.0 to 0.3 commonly used in ground-based telescopes, the RMSE of the residual wavefront is greater than 0.039λ and less than 0.041λ . This research provides an efficient and valid wavefront sensing approach for high-resolution observation with ground-based telescopes.

Keywords: Wavefront sensing; High-order turbulence; High-resolution observation; Deep learning

1. INTRODUCTION

The resolution of ground-based telescopes is severely limited by the perturbation effects caused by atmospheric turbulence, which leads to inhomogeneous changes in refractive index, affecting the propagation path of light and thereby generating significant wavefront aberrations. Adaptive optics technology, through real-time measurement by wavefront sensors, can correct these aberrations to achieve a nearly perfect diffraction-limited point spread function (PSF). Several types of pupil plane wavefront sensors (WFS) have been proposed, such as the Shack-Hartmann WFS, the plenoptic WFS, the pyramid WFS, and the curvature WFS^[1–4]. These WFS suffer, to varying degrees, from costly optical elements, stringent alignment requirements, limited dynamic range, linearity, multiple measurements, and often intensive computational reconstruction of the wavefront or the Zernike modes^[5]. A phase diversity (PD) image-based wavefront sensor (PDWFS), favored for its simple optical elements, can alleviate these downsides. Traditional PDWFS uses nonlinear optimiza-

tion algorithms to solve the unknown wavefront, but this method is prone to getting stuck in local minima^[6–9]. In recent years, PDWFSs based on deep learning have significantly improved the accuracy and speed of wavefront reconstruction^[10–21].

However, few studies have focused on deep learning wavefront sensing under high-order turbulence, despite the significant progress made in PDWFS with deep learning. Current deep learning wavefront sensing (DLWS) approaches generally follow a fixed pipeline, using Zernike polynomials to simulate the wavefront, generating the corresponding PSF, and then use a deep learning model to estimate the Zernike coefficients. For example, Paine et al. trained a convolutional neural network (CNN) by simulating a PSF to estimate Zernike coefficients for the initial wavefront aberration estimation^[9]. You et al. improved the ResNeXt network by adding a normalized attention mechanism, extending the accuracy of wavefront sensing to 67-order Zernike coefficients^[21].

Using finite Zernike polynomial modes to character-

ize the wavefront has limitations under high-order turbulence. First, undermodeling of turbulence leads to crosstalk, where some of the unestimated high-order modes appear as part of the low-order modes. Second, undersampling of turbulence brings aliasing errors, which are closely related to the sampling capability of the WFS, and will always exist even in the absence of measurement noise^[22-24]. In addition, Zernike polynomials are orthogonal only within the unit circle domain and cannot characterize discontinuous aberrations^[25]. Therefore, we use the discrete Fourier transform (DFT) method to generate an entrance pupil wavefront containing high-order turbulence to avoid the impact of crosstalk, combined with direct zonal phase output to reduce aliasing errors. At the same time, we consider discontinuous wavefront aberration caused by obstruction due to the secondary mirror and spiders when simulating the optical system model.

Here, our focus is to better understand the performance and limitations of DLWS based on PD under the background of high-order turbulence. More specifically, we study the RMSE of the estimated wavefront under the atmospheric coherence length range of 7 cm to 13 cm, with the entrance input wavefront RMSE up to 1λ , and compare the results with those corrected using Zernike modes. We also deliberately limit the simulation effect using different signal-to-noise ranges and dynamic ranges of high-order turbulence, and systematically analyze the sources of residual wavefront errors. We propose a wavefront sensing approach that can work in high-order turbulence, providing an efficient and valid solution for high-resolution observation.

2. ATMOSPHERIC MODEL AND OPTICAL IMAGING

In this study, the input data for the network model are composed of PSF pairs generated by numerical simulation, corresponding to focus PSF and defocus PSF. A simple schematic diagram of DLWS based on PD is shown in Fig. 1, where the distorted wavefront is imaged through the telescope, and the PSF of the focus channel and defocus channel serve as inputs for the network model, with the end-to-end model directly outputting the estimate wavefront. The simulation of the wavefront, containing high-order turbulence and generating the corresponding PSF, will be detailed according to the principles of optical imaging.

According to the principles of optical imaging, the PSF is defined as the Fourier transform of the generalized pupil function, described by

$$h(x, y) = |\mathcal{F}[P(x, y)]|^2, \quad (1)$$

where x and y are two-dimensional spatial coordinates, $P(x, y)$ is the generalized pupil function, and \mathcal{F} is the Fourier transform. For brevity, the spatial coordinates x and y will not be explicitly mentioned in the subsequent

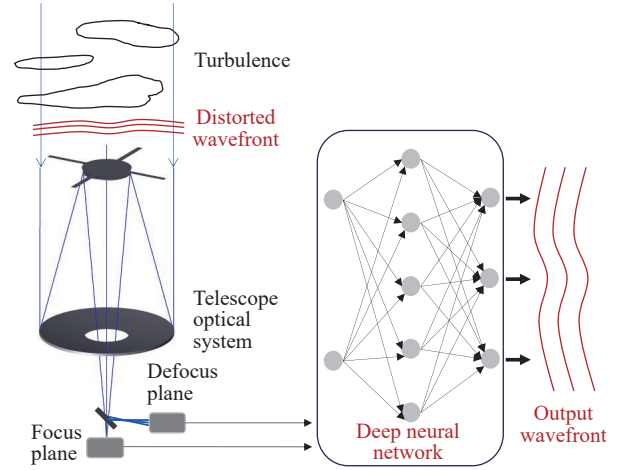


Fig. 1. Schematic diagram of the DLWS based on PD.

sections of this paper. The generalized pupil function in Equation (1) is defined as

$$P = A \cdot e^{j(\phi + \Delta\phi)}, \quad (2)$$

where ϕ is the unknown wavefront, $\Delta\phi$ is the added aberration (such as defocus), and A is the transmission function of the pupil, which is defined as

$$A = \begin{cases} 1, & \text{without obstruction} \\ 0, & \text{obstruction} \end{cases}. \quad (3)$$

In this study, we assume that the transmission rate of the obstructed area of the telescope is 0, and the transmission rate of the not-obstructed area is 1.

The unknown wavefront ϕ containing high-order turbulence is simulated according to the Kolmogorov turbulence model, where the Fried parameter, r_0 , is set to range from 7 cm to 13 cm at a wavelength of 705.8 nm^[26, 27]. The Fried parameter, r_0 , is defined as the diameter of a circular area over which the RMSE wavefront aberration, due to passage through the atmosphere, is equal to 1 radian. Typical values relevant to astronomy are in the tens of centimeters, depending on atmospheric conditions. The Kolmogorov spatial power spectrum, $\Phi(f)$, is given by

$$\Phi(f) = 0.0229 r_0^{-5/3} f^{-11/3}, \quad (4)$$

where f represents the spatial frequency. By performing DFT on the turbulence power spectrum, the wavefront containing high-order turbulence, $\phi(r)$, can be obtained as

$$\phi(r) = \mathcal{F}^{-1}[\sqrt{\Phi(f)}H(f)]. \quad (5)$$

Here, r is the spatial position in the wavefront, \mathcal{F}^{-1} is the inverse Fourier transform, and $H(f)$ is a complex filter function with unit amplitude and uniformly distributed random phase.

We have numerically compared the phase structure function of the wavefront with subharmonic compensation to the theoretical Kolmogorov phase structure function ($D_\phi^K = 6.88(\vec{r}^2/r_0)^{5/3}$). The phase structure function,

as a function of distance \vec{r} , is used to measure the correlation between different positions $\vec{x} + \vec{r}$ on the phase screen, defined as

$$D_\phi(\vec{r}) \equiv \langle |\phi(\vec{x}) - \phi(\vec{x} + \vec{r})|^2 \rangle. \quad (6)$$

By averaging the phase structure function difference between 2500 pairs of points at random from 100 generated wavefront, we verify the consistency of the DFT method with the theoretical model at all frequencies, with the results shown in Fig. 2.

The optical system parameters for our setup are referenced from the previous PD channel optical system parameters of the 1 m new vacuum solar telescope (NVST). We simulated a telescope model with an aperture of 1 m and a focal length of 25.7834 m, with a central obstruction ratio set to 0.28, and the width of the spiders to 0.02 m. The simulated detector pixel size is $9.2 \mu\text{m} \times 9.2 \mu\text{m}$, and ensures a sampling rate of 2 for the PSF. The PSF corre-

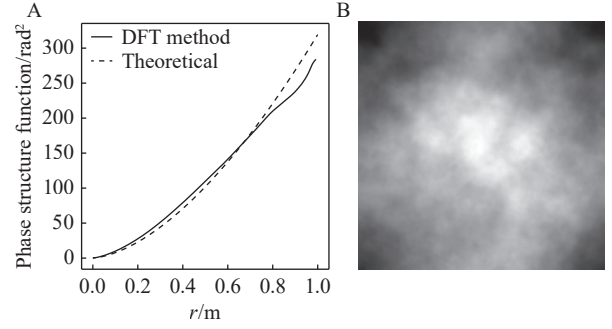


Fig. 2. (A) Phase structure function and (B) phase screen sample. The dimension of the phase screen is 1 m×1 m, the number of sampling points is 256×256 and the Fried parameter is 10 cm.

sponding to the wavefront containing high-order turbulence can be obtained through Equation (1). In PD, it is typical to add a defocus of 1λ (peak-to-valley) to $\Delta\phi$. Fig. 3 shows a minibatch of focus and defocus PSF pairs.

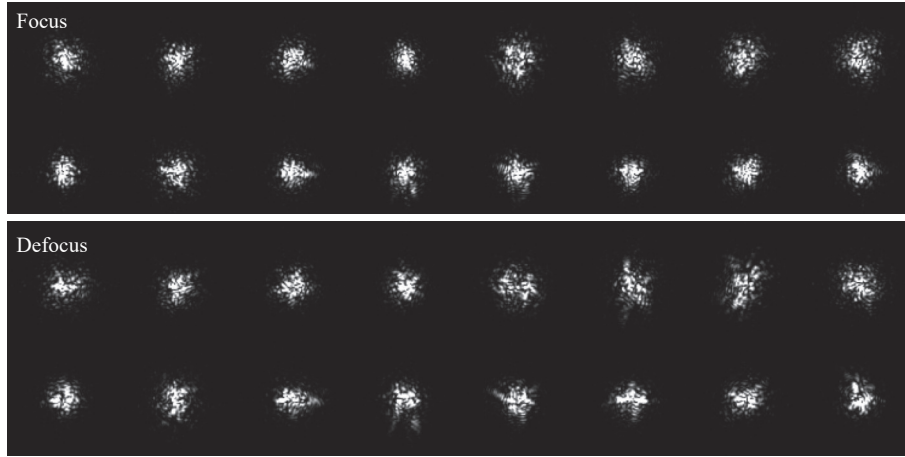


Fig. 3. Example of a minibatch of input PSF used to training.

3. NETWORK ARCHITECTURES AND TRAINING

3.1. Squeeze-and-Excitation Model

In the field of DLWS, mapping intensity images to the wavefront phase is the goal, and feature extraction plays a crucial role in this. The CNN extracts new feature maps from input feature maps using convolutional kernels, and as the network depth increases, the network can learn more complex features, which may lead to better sensing results. In existing research, DLWS models have been constructed using network structures such as Inception and Xception[28, 29]. However, as the network depth increases, the vanishing gradient problem may lead to models failing to converge during training. ResNet solved this problem with skip connections, but to further improve accuracy, the network still needs to be deepened[30].

Inspired by the Squeeze-and-Excitation (SE) module, we aimed for the model to learn the importance of fea-

tures from different channels, strengthening significant channels and weakening less significant ones[31]. As shown in Fig. 4 and Fig. 5, we incorporated a SE module in the residual block (SE blocks) and embedding it into network model, only 16 layers of SE blocks are required to accurately map PSF images to the wavefront. Details of each block are described in Table 1.

3.2. Optimization

We use the entrance pupil wavefront ϕ corresponding to the PSF, as the label for the training data, and optimize the neural network by minimizing the difference between the output estimated wavefront, $\hat{\phi}$, and the entrance pupil wavefront, ϕ . The loss function is defined by the relationship

$$\text{loss} = \sum |\phi - \hat{\phi}|^2. \quad (7)$$

Training uses the Adam optimizer with an initial learning rate of 0.001. The model is first trained on PSF

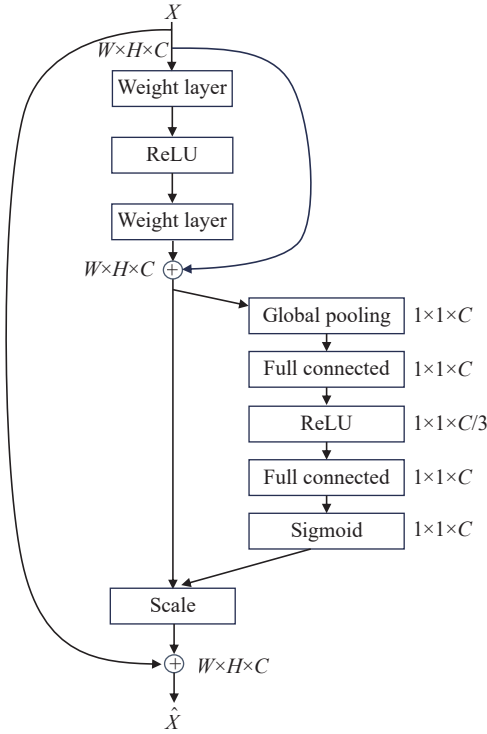


Fig. 4. Architecture of the SE module into a residual block. SE module enable the network to perform dynamic channel-wise feature recalibration.

datasets, generated without obstruction, for 3600 epochs. The model is then transferred to PSF datasets with central obstruction and spider obstruction for an additional 600 epochs of training, to adapt to discontinuous wavefront aberrations.

In the final training stage, we add simulated noise, including detector readout noise and Poisson noise, with noise intensity randomly distributed between 1% and 2%

of the PSF intensity. After training on PSF datasets with added noise for 600 epochs, the model demonstrates robustness to noise, with the final residual wavefront RMSE being 0.03λ .

4. RESULTS AND ANALYSIS

4.1. Performance of DLWS

We test the accuracy of DLWS on a test dataset different from the training and valid datasets. The test dataset includes wavefront samples containing high-order turbulence generated in the same way as the training dataset, with the Fried parameter, r_0 , ranging between 7 cm and 13 cm. The test dataset contains 4 000 wavefront samples for each r_0 value to ensure that the test results are statistically significant.

First, we calculated the mean and standard deviation of the residual wavefront RMSE for the test dataset and compared it with the Marechal criterion. According to the Marechal criterion, if the wavefront RMSE error is 13.4 times smaller than its wavelength (0.075λ), it is considered diffraction-limited. Fig. 6 shows the curve of residual wavefront RMSE as a function of r_0 , with the black solid line showing the mean residual wavefront RMSE, the shaded area showing the standard deviation range of the residual wavefront RMSE, and the red dashed line showing the Marechal criterion. For all r_0 values, the mean residual wavefront RMSE of the test dataset is within the diffraction-limited wavefront RMSE of 0.075λ . The magnitude of the residual wavefront RMSE is negatively correlated with r_0 , and as r_0 increases, the residual wavefront RMSE tends to flatten.

Conan, based on Noll's rigorous analysis, provided an approximate expression for the residual wavefront RMSE

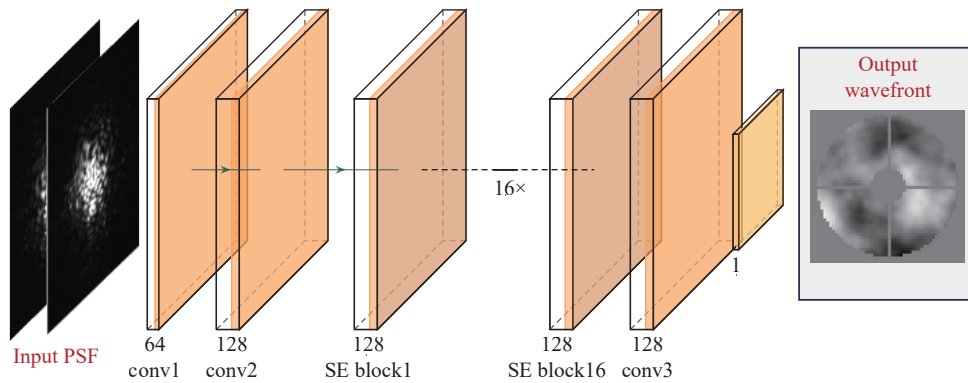


Fig. 5. Architecture of DLWS.

Table 1. Details of each block of DLWS

Layer	Type	Kernel size	Padding	Input shape	Output shape
$L_{1,1}$	conv2D	3	1	$100 \times 100 \times 2$	$100 \times 100 \times 64$
$L_{2,1}$	conv2D	3	1	$100 \times 100 \times 64$	$100 \times 100 \times 128$
$L_{3,1} - L_{19,1}$	(SE) block				
$L_{20,1}$	conv2D	3	1	$100 \times 100 \times 128$	$100 \times 100 \times 128$
$L_{21,1}$	conv2D	3	1	$100 \times 100 \times 128$	$50 \times 50 \times 1$

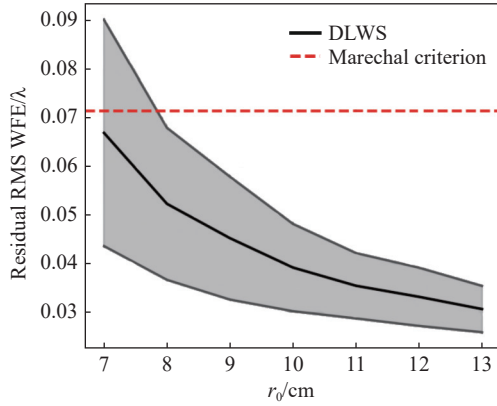


Fig. 6. Residual wavefront RMS for different r_0 . The shaded area is the standard deviation range of the residual wavefront RMS, the dashed red line is $1/13.4\lambda$, i.e. the Marechal criterion.

after Zernike mode correction in adaptive optics^[25, 32] as

$$\sigma^2 = 0.2944N_m^{-\sqrt{3}/2} \left(\frac{D}{r_0} \right)^{5/3}, \quad (8)$$

where N_m is the number of Zernike modes required for correction, D is telescope diameter, and σ^2 is residual wavefront variance after corrected. Table 2 provides a comparison of the results, with the first column giving the Fried parameter, the second column giving the corresponding residual wavefront RMSE, the third column showing the diffraction-limited wavefront RMSE, and the fourth column being the number of Zernike modes calculated using Equation (5). The results prove that DLWS, which uses direct regional phase representation to include high-order turbulence patterns, reaches turbulence patterns at a higher order than the finite-order Zernike modes. To our knowledge, this result is the highest accuracy reported to date for deep learning wavefront probing methods based on the PD principle.

Table 2. Zernike modes, N_m , required for correction

r_0/cm	Residual RMS WFE	Diffraction-limited	N_m
7	0.067λ		301
8	0.052λ		411
9	0.045λ		458
10	0.039λ	0.075λ	521
11	0.035λ		546
12	0.033λ		539
13	0.031λ		554

We use samples from two different r_0 to illustrate how an increase in r_0 affects the PSF and wavefront reconstruction. Fig. 7(a) compares the entrance pupil wavefront and DLWS output wavefront for r_0 values of 7 cm and 10 cm. Observing from a single wavefront perspective, the r_0 value of 7 cm wavefront contains more high-order turbulence than the r_0 value of 10 cm wavefront. Consequently, the PSF shapes before correction in Fig. 7(b) (left) are more complex. Fig. 7(b) (bottom) quantitatively

describes the PSF before and after correction using encircle energy. The green solid line is the encircle energy distribution of the PSF before correction, the orange solid line is the encircle energy distribution of the PSF after correction, and the blue solid line is the encircle energy distribution of the diffraction-limited PSF. From the encircle energy, it can be observed that when the entrance pupil wavefront contains more high-order turbulence, and the corresponding PSF has a broader energy distribution with a more complex shape. The error between the encircle energy of the corrected PSF and the diffraction-limited PSF is significantly reduced when the high-order turbulence decrease.

4.2. Limitations of DLWS

In the DLWS results, the residual wavefront RMSE increases with increasing high-order turbulence. There are various factors affecting accuracy, such as fitting errors or dynamic range limitations.

We first analyze the fitting error in wavefront reconstruction under high-order turbulence. As shown in Fig. 8, the wavefront containing high-order turbulence is projected onto first 465-order Zernike modes, where the gray solid line is the entrance pupil wavefront, the red solid line is the output wavefront, and the orange solid line is the residual wavefront. Fig. 8 shows that there is no obvious correlation between the residual wavefront RMSE and Zernike modes. DLWS uses direct zonal phase reconstruction, which fits well with different orders of Zernike mode. Because the residual wavefront RMSE is the sum of errors from various Zernike modes, the accumulated fitting errors of numerous high-order Zernike modes may be the main source of residual wavefront RMSE.

Fitting errors of high-order turbulence affect the residual wavefront RMSE. This is shown in Fig. 9, where from top to bottom are the entrance pupil wavefront, the corresponding PSF, the wavefront output by DLWS, and the residual wavefront. The first column is the unmodified wavefront containing high-order turbulence, with a wavefront RMSE of 0.782λ . In the second column, after removing some high-order turbulence, the wavefront with small amount of high-order turbulence has an RMSE of 0.769λ . The fitting error, especially at the edge, is significantly reduced after removing some high-order turbulence, with the residual error RMSE decreasing from 0.545λ to 0.312λ . Fig. 10 shows a scatter plot comparison between the entrance pupil wavefront and the output wavefront, where the blue dots are the unmodified wavefront containing high-order turbulence, and the red dots are the wavefronts with small amount of high-order turbulence. The blue dots have a noticeably higher fitting error compared with the red dots.

DLWS is also limited by the dynamic range of the entrance pupil wavefront RMSE, where we analyzed the accuracy of the output wavefront for entrance pupil wavefront of different aberration levels. As shown in Fig. 11, we trained three models with three datasets containing dif-

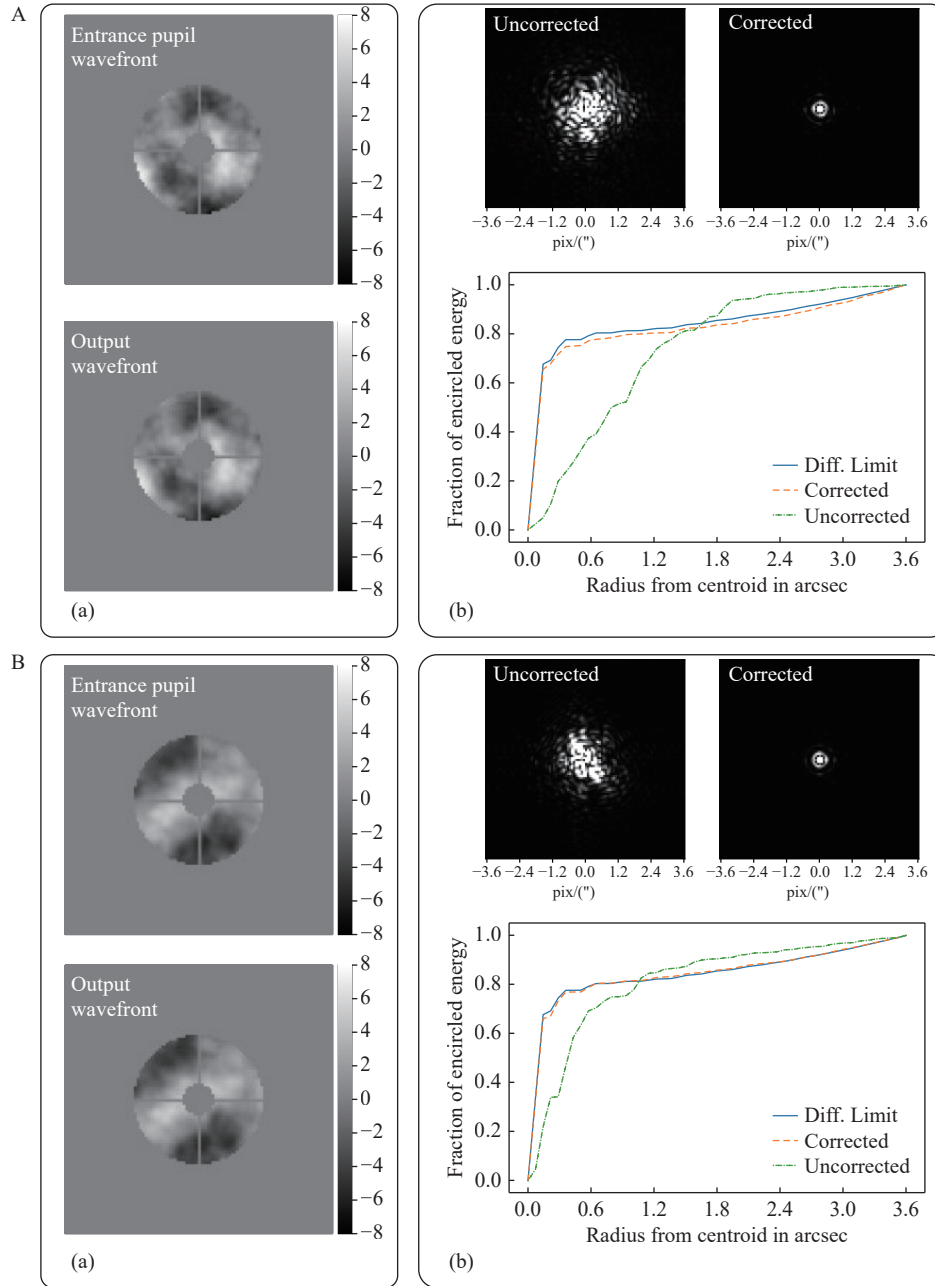


Fig. 7. The PSF and wavefront reconstruction affected by r_0 change, $r_0=7$ cm (A) and $r_0=10$ cm (B). (a) shows the comparison of the entrance pupil wavefront and output wavefront, (b)(upper) is the comparison of uncorrected focus PSF and corrected focus PSF, and (b) (lower) is the encircled energy comparison of uncorrected focus PSF, corrected focus PSF, and diffraction-limited PSF.

ferent aberration levels, where the blue solid line is trained with a dataset ranging from 0.5λ to 1.0λ , the orange solid line is trained with a training dataset ranging from 0.1λ to 0.5λ , and the green solid line is trained with a training dataset ranging from 0.1λ to 1.0λ . These models were tested on 11 different test datasets with RMSE ranging from 0.1λ to 1.5λ . We find that at low aberration levels, the accuracy of the three models is similar, and at higher aberration levels, the model trained with the 0.1λ to 0.5λ range dataset rapidly decreases in accuracy. The behavior at high aberration levels results from the way in which we generate the dataset, with high aberration levels containing more high-order turbulence. The

model trained with the low aberration level dataset has a small dynamic range, hence the fitting error increases. The two model trained with high aberration levels have similar accuracy at different aberration levels because the training set implicitly includes low-order turbulence.

5. DISCUSSION

5.1. Robustness to Changing Signal-to-Noise Ratio (SNR)

In this section, we explore the robustness of the DLWS model at different SNR levels. The DLWS was

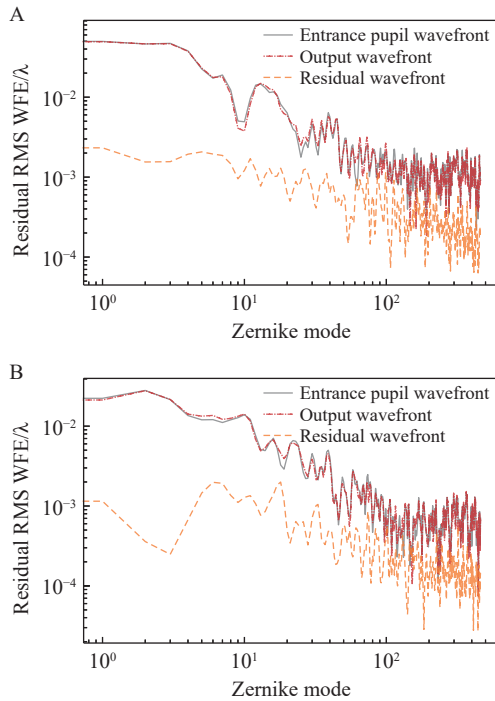


Fig. 8. Modal RMS wavefront error for two different levels of aberrations: $r_0=7$ cm (A) and $r_0=10$ cm (B). Residual wavefront RMS are largely affected by high-order modes, which are the accumulated fitting errors of numerous high-order Zernike modes.

trained at a SNR range of 1% – 2% and then tested over a broader range of SNR during testing. Fig. 12 shows the curve of the test set with $r_0=10$ cm as the SNR varies. The black solid line is the average value of the residual wavefront RMSE, the gray area is the standard deviation range of the residual wavefront RMSE, the red dashed line is the diffraction limit, and the orange area is the SNR range of the training dataset. Within the range of the training dataset, the residual wavefront RMSE is essentially maintained near the diffraction limit. As the SNR increases, the RMSE gradually tends towards a fixed value. When the SNR is increased from 1% to 5%, the residual wavefront RMSE rises significantly, which is related to the SNR range of the training set. If robustness over a wide range of SNR levels is required, the SNR range of the training set can be adjusted. The distribution of the SNR in the training dataset should be established based on the expected magnitude range, desired accuracy, and observational variations. Alternatively, multiple models can be used, and different models can be selected based on different magnitudes.

5.2. Robustness to Changing Central Obstruction Ratio (CO)

In this section, we explore the robustness of the DLWS model at different central obstruction levels. Central obstruction affects the diffraction pattern in the optical system. As the CO increases, the diffraction rings become more pronounced, leading to a decrease in perfor-

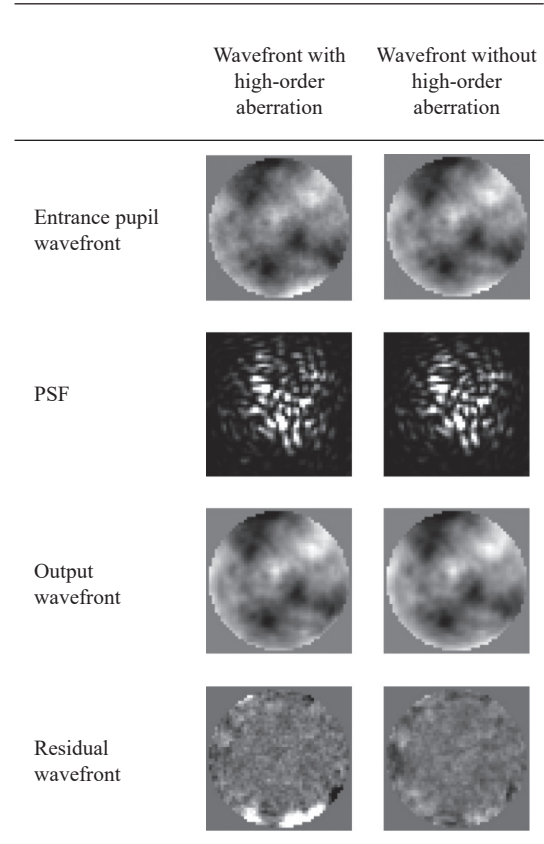


Fig. 9. Comparison of entrance pupil wavefront, output wavefront, and residual wavefront with a large amount of high-order turbulence and a small amount of high-order turbulence.

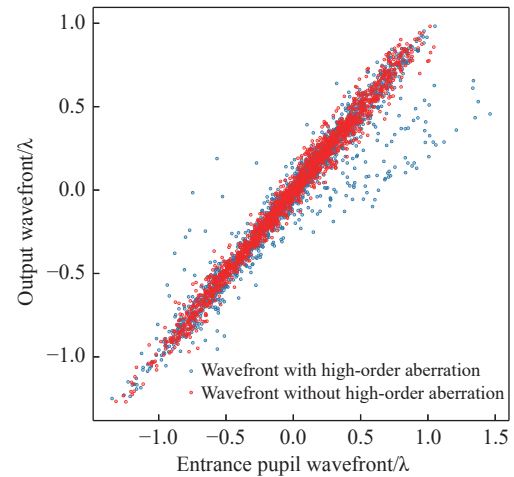


Fig. 10. Scatter plot comparison between the entrance pupil wavefront and the output wavefront.

mance of the optical system for high-resolution imaging. Generally, the larger the obstruction ratio, the lower the light transmission efficiency of the system. Fig. 13 shows the curve of the test dataset with $r_0=10$ cm as the CO varies, where the black solid line is the average value of the residual wavefront RMSE, and the gray area is the standard deviation range of the residual wavefront RMSE. The diffraction-limited PSF shapes for CO = 0, 0.28, and

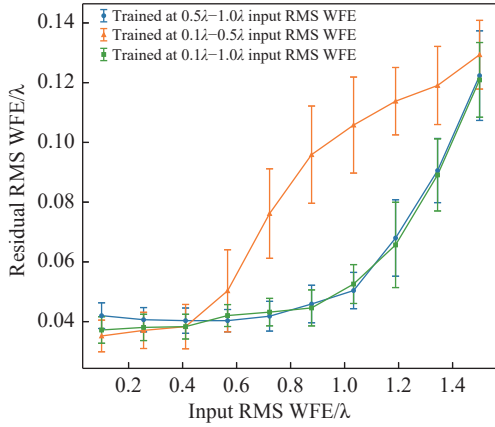


Fig. 11. Effect of the dynamic range of the entrance pupil wavefront RMS on the residual wavefront RMS.

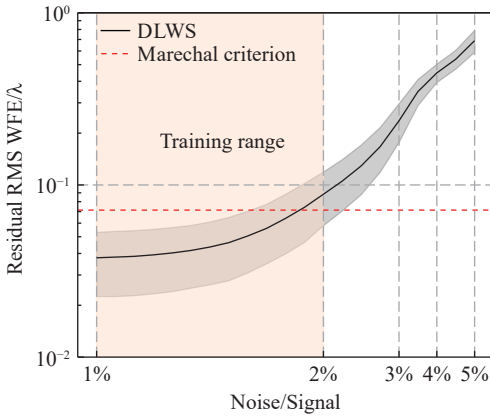


Fig. 12. The curve of the test dataset with $r_0=10$ cm as the SNR varies.

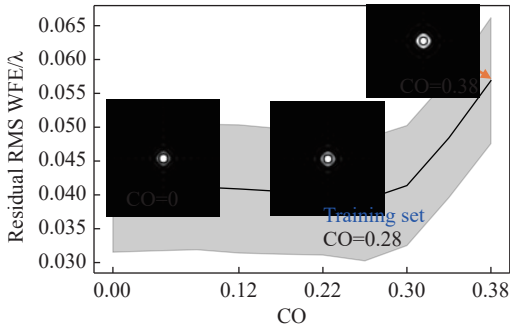


Fig. 13. The curve of the test dataset with $r_0=10$ cm as the CO varies.

0.38 are also shown individually. Within the 0.28 CO range, the change in residual wavefront RMSE is relatively flat, and the smallest residual wavefront RMSE occurs at the specific 0.28 CO point used in the training dataset. As the CO increases, the phase in local areas of the wavefront changes significantly. This high-frequency error directly affects image resolution, causing the residual wavefront RMSE to increase significantly. Typically, the CO of a ground-based telescope is within 0.3, and DLWS is robust over a wide range of CO up to 0.3. When the CO is greater than 0.3, because of diffraction

effects, training is required with datasets specific to the optical system.

6. CONCLUSION

We have developed an end-to-end approach for the direct sensing of the entrance pupil wavefront from phase diversity PSF images. In this context, we consider the effect of the high-order turbulence and discontinuous aberration caused by optical obstructions encountered in real-world ground-based telescope observations. Our simulations show that DLWS can correct the entrance pupil wavefront to near the diffraction limit under conditions of high-order turbulence, within the atmospheric coherence length range of 7 cm to 13 cm.

We also prove that the residual wavefront error of DLWS, which uses a direct zonal phase output method, mainly originates from fitting errors caused by the high-order turbulence. The dynamic range of DLWS is closely related to the dynamic range of the training dataset; training with an entrance pupil wavefront dynamic range (0.1λ to 1λ) that matches real-world observations enables DLWS to work on pupil wavefronts outside the dynamic range (1.0λ to 1.5λ). Our DLWS model, trained with a central obstruction ratio of 0.28, shows that within the commonly used central obstruction ratios for ground-based telescopes, ranging from 0.0 to 0.3, the residual wavefront RMSE is greater than 0.039λ and less than 0.041λ . However, the wavefront reconstruction accuracy of DLWS is susceptible to the influence of noise intensity. The test dataset residual wavefront RMSE is less than 0.075λ within a noise intensity range of 1% to 1.9%. As the noise intensity increases, the wavefront reconstruction accuracy significantly decreases. Therefore, to maintain robustness over a wide range of noise levels, the training strategy needs to be adjusted accordingly.

Although our study validates the reliability of DLWS under high-order turbulence, real-world data are more complex than the images we simulate, with limited wavelength range, different detector noise, and imperfect optical alignment. Some of these effects can be predicted and simulated, such as wavelength range, which will filter out some high spatial frequency information, but others may not be accurately simulated. Although the optical system model can be improved to generate more realistic labeled datasets, experimental data are still ultimately needed for accurate training. A key challenge for future work to address is how to utilize a limited set of experimental data.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (NSFC) (U2031140).

AUTHOR CONTRIBUTIONS

Dongming Liu conceived the ideas, designed and imple-

mented the study, and wrote the paper. Hui Liu conceived the ideas and designed the validated method. Zhenyu Jin conceived methodology. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

Zhenyu Jin is the associate editor-in-chief for *Astronomical Techniques and Instruments* and he was not involved in the editorial review or the decision to publish this article. The authors declare no competing interests.

REFERENCES

- [1] Platt, B. C., Shack, R. 2001. History and principles of Shack-Hartmann wavefront sensing. *Journal of Refractive Surgery*, **17**(5): S573–S577.
- [2] Wu, C. S., Ko, J., Davis, C. C. 2015. Determining the phase and amplitude distortion of a wavefront using a plenoptic sensor. *Journal of the Optical Society of America A*, **32**(5): 964–978.
- [3] Ragazzoni, R. 1996. Pupil plane wavefront sensing with an oscillating prism. *Journal of Modern Optics*, **43**(2): 289–293.
- [4] Roddier, F. 1988. Curvature sensing and compensation: a new concept in adaptive optics. *Applied Optics*, **27**(7): 1223–1225.
- [5] Carbillet, M., Ferrari, A., Aime, C., et al. 2006. Wavefront sensing: from historical roots to the state-of-the-art. *European Astronomical Society Publications Series*, **22**: 165–185.
- [6] Gonsalves, R. A., Chidlaw, R. 1979. Wavefront sensing by phase retrieval. In *Proceedings of SPIE Applications of Digital Image Processing III*. 207: 32–39.
- [7] Gonsalves, R. A. 1982. Phase retrieval and diversity in adaptive optics. *Optical Engineering*, **21**(5): 829–832.
- [8] Paxman, R. G., Fienup, J. R. 1988. Optical misalignment sensing and image reconstruction using phase diversity. *Journal of the Optical Society of America A*, **5**(6): 914–923.
- [9] Paxman, R. G., Schulz, T. J., Fienup, J. R. 1992. Joint estimation of object and aberrations by using phase diversity. *Journal of the Optical Society of America A*, **9**(7): 1072–1085.
- [10] Ju, G. H., Qi, X., Ma, H. C., et al. 2018. Feature-based phase retrieval wavefront sensing approach using machine learning. *Optics Express*, **26**(24): 31767–31783.
- [11] Swanson, R., Lamb, M., Correia, C., et al. 2018. Wavefront reconstruction and prediction with convolutional neural networks. In *Proceeding of SPIE Adaptive Optics Systems*. 10703: 481–490.
- [12] Ma, H. M., Liu, H. Q., Qiao, Y., et al. 2019. Numerical study of adaptive optics compensation based on convolutional neural networks. *Optics Communications*, **433**: 283–289.
- [13] Nishizaki, Y., Valdivia, M., Horisaki, R., et al. 2019. Deep learning wavefront sensing. *Optics Express*, **27**(1): 240–251.
- [14] Xin, Q., Ju, G. H., Zhang, C. Y., et al. 2019. Object-independent image-based wavefront sensing approach using phase diversity images and deep learning. *Optics Express*, **27**(18): 26102–26119.
- [15] Guo, H. Y., Xu, Y. J., Li, Q., et al. 2019. Improved machine learning approach for wavefront sensing. *Sensors*, **19**(16): 3533.
- [16] Wu, Y., Guo, Y. M., Bao, H., et al. 2020. Sub-millisecond phase retrieval for phase-diversity wavefront sensor. *Sensors*, **20**(17): 4877.
- [17] Andersen, T., Owner-Petersen, M., Enmark, A. 2020. Image-based wavefront sensing for astronomy using neural networks. *Journal of Astronomical Telescopes, Instruments, and Systems*, **6**(3): 034002.
- [18] Vera, E., Guzmán, F., Weinberger, C. 2021. Boosting the deep learning wavefront sensor for real-time applications. *Applied Optics*, **60**(10): B119–B124.
- [19] Li, Y. S., Yue, D., He, Y. H. 2022. Prediction of wavefront distortion for wavefront sensorless adaptive optics based on deep learning. *Applied Optics*, **61**(14): 4168–4176.
- [20] Zhou, Z. S., Fu, Q., Zhang, J. Q., et al. 2023. Generalization of learned Fourier-based phase-diversity wavefront sensing. *Optics Express*, **31**(7): 11729–11744.
- [21] You, J., Gu, J. L., Du, Y. L., et al. 2023. Atmospheric turbulence aberration correction based on deep learning wavefront sensing. *Sensors*, **23**(22): 9159.
- [22] Herrmann, J. 1981. Cross coupling and aliasing in modal wave-front estimation. *Journal of the Optical Society of America*, **71**(8): 989–992.
- [23] Southwell, W. H. 1983. What's wrong with cross coupling in modal wave-front estimation? In *Proceeding of 26th Annual Technical Symposium*. SPIE, 365: 97–104.
- [24] Quirós-Pacheco, F., Conan, J. M., Petit, C. 2010. Generalized aliasing and its implications in modal gain optimization for multi-conjugate adaptive optics. *Journal of the Optical Society of America A*, **27**(11): A182–A200.
- [25] Noll, R. J. 1976. Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, **66**(3): 207–211.
- [26] Kolmogorov, A. N. 1941. The local structure of turbulence in incompressible viscous fluid for very large Reynolds. *Doklady Akademii Nauk SSSR*, **30**: 301–305.
- [27] Fried, D. L. 1966. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. *Journal of the Optical Society of America A*, **56**(10): 1372–1379.
- [28] Szegedy, C., Vanhoucke, V., Ioffe, S., et al. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [29] Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*.
- [30] He, K. M., Zhang, X. Y., Ren, S. Q., et al. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [31] Hu, J., Shen, L., Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [32] Conan, R. 2008. Mean-square residual error of a wavefront after propagation through atmospheric turbulence and after correction with Zernike polynomials. *Journal of the Optical Society of America A*, **25**(2): 526–536.